# LANGUAGE IDENTIFIER



- TEAM MEMBERS :-
- 1) Srinivas Sirigiri (SE22UARI167)
- 2) Perubogu Akhil Raj (SE22UAI196)
- 3) Pyata Suhrudh Reddy (SE22UARI194)

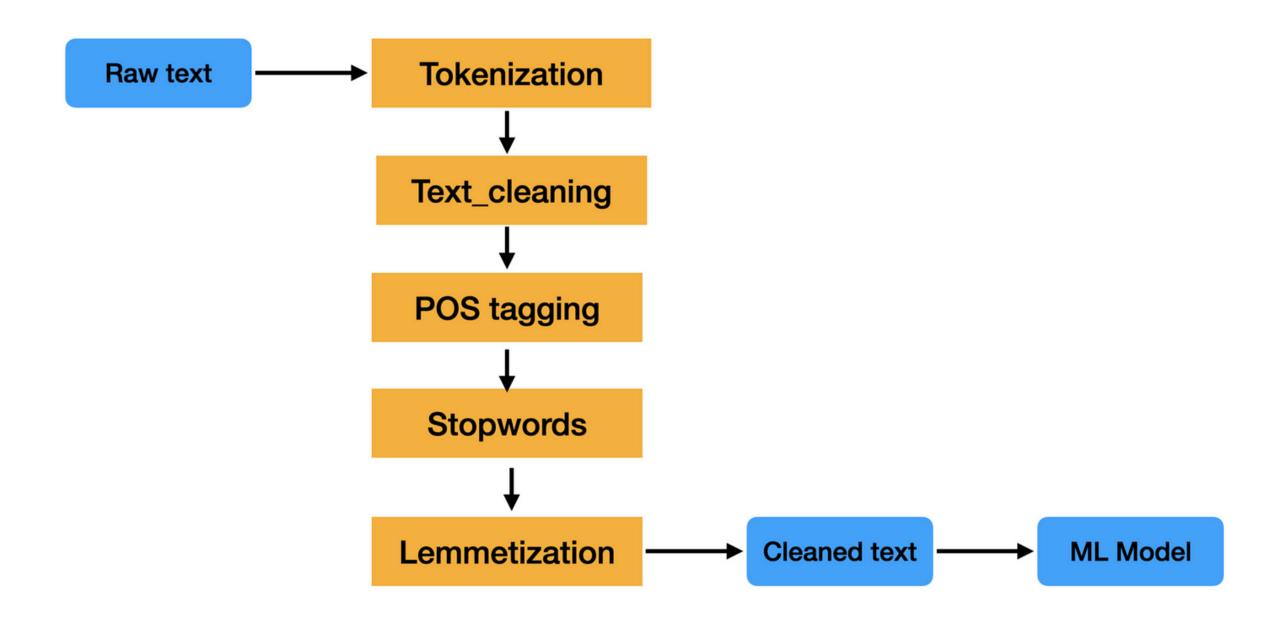
#### Language Identifier :-

Language identification, also known as language guessing, is a process in natural language processing (NLP) that determines which natural language a given piece of text is written in.

#### •Why?

- 1) Enhancing User Experience:-customer support systems.
- 2) Data Processing Efficiency:- For NLP models.
- 3) Improving Accuracy in Applications:-speech recognition.
- 4) Spam Filtering and Security:-combat spam.

- Where is it used?
- 1) Translation Services.
- 2) Social Media Monitoring for sentiment analysis.
- 3) Search Engines to improve accuracy of search results.
- 4) Speech Recognition Systems.



- How is NLP a part of Language Identification?
- 1) NLP techniques are utilized to accurately identify the language of input text.
- 2) Autocorrection and Predictive Text.
- 3) Search Engines to improve accuracy of search results where NLP helps identifying the language and increases accuracy of results.
- 4) Social Media tools:- to identify languages used in posts or messages.

#### First Moves Towards The Goal

#### Languages

 Select what languages we are about to consider for the language identifier

#### **Dataset**

 Gather a diverse dataset that includes text samples in multiple languages

# Data **Preprocessing**

- Tokenization
- Stopword Removal
- Normalization

# Thank You.